End-to-end Stochastic Scheduling of Scalable Video Over Time Varying Channels

Nesrine Changuel, Nicholas Mastronarde, Mihaela van Der Schaar, Bessem Sayadi, Michel Kieffer

▶ **To cite this version:**

Nesrine Changuel, Nicholas Mastronarde, Mihaela van Der Schaar, Bessem Sayadi, Michel Kieffer. End-to-end Stochastic Scheduling of Scalable Video Over Time Varying Channels. MM'10: International conference on Multimedia, Oct 2010, Florence, Italy. pp.731-734, 10.1145/1873951.1874064 . hal-00549244

HAL Id: hal-00549244
https://hal.science/hal-00549244

Submitted on 21 Dec 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

End-to-end Stochastic Scheduling of Scalable Video Over Time-Varying Channels

Nesrine Changuel
Alcatel-Lucent - Bell-Labs France
91620 Nozay, France
nesrine.changuel@alcatel-lucent.com

Nicholas Mastronarde
UCLA, 420 Westwood Plaza
Los Angeles, CA 90095-1594, USA
nhmastro@gmail.com

Mihaela van derSchaar
UCLA, 420 Westwood Plaza
Los Angeles, CA 90095-1594, USA
mihaelaucla@gmail.com

Bessem Sayadi
Alcatel-Lucent - Bell-Labs France
91620 Nozay, France
bessem.sayadi@alcatel-lucent.com

Michel Kieffer*
LTCI, CNRS-Télécom ParisTech
75013 Paris, France
kieffer@lss.supelec.fr

ABSTRACT

This paper addresses the problem of video on demand delivery over a time-varying wireless channel. Packet scheduling and buffer management are jointly considered for scalable video transmission to adapt to the changing channel conditions. A proxy-based filtering algorithm among scalable layers is considered to maximize the decoded video quality at the receiver side while keeping a minimum playback margin. This problem is cast in the context of Markov Decision Processes which allows the design of foresighted policies maximizing some long-term reward. Experimental results illustrate the benefit of this approach compared to a short-term policy in term of average PSNR improvement.

Keywords

Buffer control, Markov Decision Process, Scalable video coding, Video scheduling

Categories and Subject Descriptors

C.2 [**Computer communication network**]: Network Architecture and Design—*Wireless communication, Network management, Client/server*

General Terms

Algorithms, Experimentation

*M. Kieffer is on leave from L2S, Univ Paris-Sud - CNRS - SUPELEC, 91192 Gif-sur-Yvette, France.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
MM'10, October 25–29, 2010, Firenze, Italy.
Copyright 2010 ACM 978-1-60558-933-6/10/10 ...$10.00.

1. INTRODUCTION

The rapid growth in wireless networks allows a higher diversity of services provided to users. Thus, broadcasting, mobile television, or video-on-demand are expected to develop widely in the near future. Even if wireless resources expected to grow significantly for large-display mobile receivers, one major concern is to maximize the received video quality. This paper focuses on Quality of Service (QoS) optimization in the context of video streaming applications to mobile users while considering channel and buffering constraints. The time-varying characteristics of the source and of the channel make the end-to-end optimization of such streaming chain quite difficult.

Several components in the streaming system may be optimized, for example, the media server, intermediate buffers, the channel resource scheduler, receiver buffers, *etc.* [9].

A first class of techniques used to solve this problem considers a deterministic optimization framework. The work in [2] focuses on delivery delay satisfaction of real-time non scalable video contents. To solve this problem accurate models of the rate-distortion (R-D) characteristics of the source are required. In [4], a scheme is proposed to avoid underflow state of the playback buffer while smoothing out the variations of the encoding rate. An adaptive media playout method is proposed in [5] to limit the data buffering at the client and to reduce the display delay in time-varying channel conditions. In [6], the problem is extended to a multi-users scenario.

A second class of techniques consider Markov models of the source, the contents of the buffers, or the channel to perform a stochastic control of the streaming system. In [1], the quantization parameters are optimized based on source and buffers Markov models for non scalable video. The problem is cast into the framework of Markov Decision Processes (MDP) [11]. Scalable video coders are considered in [3], where the impact of error-concealment which may be performed at receiver side is explicitly taken into account when searching for some optimal long-term control policy. Later, [7] proposes a framework for the design of autonomous layered video coders by optimizing each layer accounting for

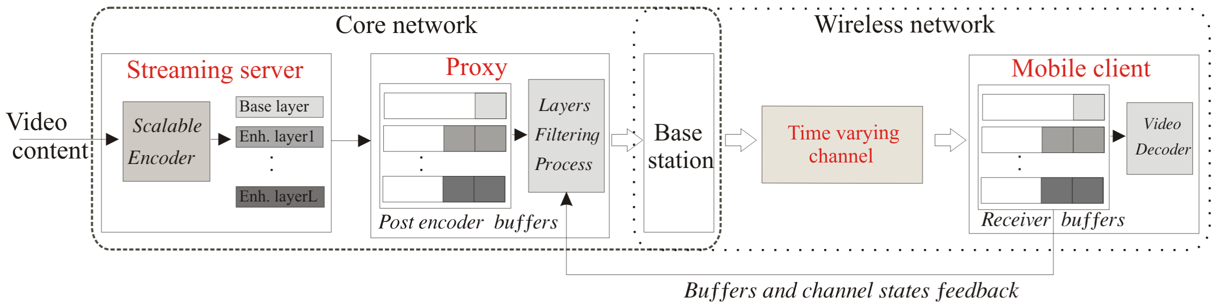

Figure 1: End-to-end video streaming scheme

some information provided by the other layers and for the impact of the parameters chosen by a given layer on the other layers.

In this paper, we consider the problem of joint packet scheduling for SNR scalable video coders. Encoder and receiver buffers management is also considered with a stochastic control technique over time-varying channels. A proxy-based filtering algorithm among scalable layers is proposed to maximize packet transmission while considering priority level among layers, see Section 2. Buffers at encoder and receiver side are considered for each layer to ensure minimum playback margin and maintain satisfying video quality. This problem is cast in the context of Markov Decision Processes (MDP) in Section 3. This formalism allows to derive a foresighted control policy maximizing some long-term discounted sum of rewards. Experimental results, detailed in Section 4 illustrate the benefits of this approach compared to a short-term (myopic) policy. The impact of information concerning the level of the receiver buffers on the global scheduling performance is mentioned.

2. STREAMING SYSTEM

The considered streaming system is illustrated Figure 1. The core network consists of a streaming server, hosting a scalable video coder, a *proxy*, and a base station, which is the front-end of the wireless part of the system. Packets are transmitted through a wireless channel and received by a mobile client. Information on the level of the receiver buffers may be sent back to the proxy.

2.1 System description

In the streaming server, the video sequence is segmented into frames and encoded into a base layer and a set of $L-1$ enhancement layers. The base layer and its corresponding enhancement layers form an access unit (AU). An AU is generated with a constant period of time Δt and is identified by its temporal index t.

Scalable Video Coding (SVC) [10] supports usually three types of scalability: temporal, spatial, and quality (SNR) scalability. SNR scalability is classified into Coarse, Medium, and Fine-Grain quality Scalable coding (CGS, MGS, FGS). Here, only MGS scalability is considered, since it is well-suited in an unicast scenario. The encoding parameters (quantization steps, frame rate, etc.) are controlled by the streaming server, independently of the remainder of the chain.

In order to minimize the drift due to lost layers, a *base layer only* control scheme [10] for the encoder is considered. Each SNR layer of each encoded frame is packetized into a single Network Abstraction Layer (NAL) unit, which itself is encapsulated into a Real-time Transport Protocol (RTP) packet. These packets are fed via the an over-provisioned core network (assumed lossless), to the L post-encoder buffers of the proxy. Considering one buffer per layer facilitates differentiation of the actions applied to each layer.

For each layer, the proxy has to decide to send packets, to wait, or to drop packets (*layer filtering process*). Constraints on the available bandwidth have to be satisfied. For that purpose, the proxy may exploit some feedback from the mobile client to estimate the channel conditions. In order to limit the delay between transmission and feedback information, the proxy is placed at the boundary of the wireless network, close to the base station. Here, delays resulting from buffering at MAC layer and transmission are neglected.

The receiver hosts the video decoder and one buffer per scalability layer. The levels of the receiver buffers and the state of the channel are fed back (with no delay nor error) to the proxy with a period Δt. For both post-encoder and receiver buffers, when buffers reach fullness, packets in the queue have to be dropped in a Head-Of-Line (HOL) order, *i.e.*, packet which resides longest in the buffer are dropped first. At each time t, the decoder builds AUs from the packets available in the receiver buffers, which are then decoded. Outdated packets are dropped, without being decoded.

2.2 System constraints

This paper focuses on the design of an efficient layer filtering process done by the proxy in such a way that the quality of the decoded video is maximized while satisfying the following constraints: (i) the transmission rate has to be below and as close as possible to the rate allowed by the channel; (ii) the level of the post-encoder buffers should avoid under and overflow; (iii) the receiver buffers should provide some playback margin to be robust against temporary unavailability of the channel.

3. SYSTEM MODEL

In this section, the problem of designing an optimal scheduling policy of L SNR scalable layers over a wireless channel is translated in the framework of discrete-time MDP [11].

An MDP is a 4-tuple $(\mathcal{S}, \mathcal{A}, T, R)$, where $\mathcal{S}$ is the set of states of the system, $\mathcal{A}$ is the set of actions, $T(s, s', a)$ determines the transition probability from $s \in \mathcal{S}$ at time $t-1$ to $s' \in \mathcal{S}$ at time t, when the action $a \in \mathcal{A}$ is applied to the system. Finally $R(s, s', a)$ indicates the immediate reward received after transition from s to s' with transition when a is applied on the system.

Designing an optimal scheduler for the proxy consists thus in determining an optimal policy $\pi(s)$, $s \in \mathcal{S}$. Such policy may be obtained using, *e.g.*, classical value iteration technique, see [11]. This requires all components of the tuple to be identified for the system considered in Section 2.

3.1 States

The considered states of the system are the states of the channel $h_t \in \mathcal{H} = 0, 1$, the levels of the post-encoding buffers hosted by the proxy s_l^e, $l = 1 \dots L$, and the levels of the receiver buffers s_l^r, $l = 1 \dots L$. More states could be considered (type of picture) to get a more accurate control process at the price of an increased complexity.

3.1.1 Channel model

The behavior of the channel is described by a two-state Markov model, to simulate the bursty nature of an error-prone wireless channel. The channel state h_t represents the channel conditions, assumed constant, between time $t-1$ and t. In the *good* state ($h_t = 1$), at most R_c bits/s may be transmitted. In the *bad* state ($h_t = 0$), the channel is unable to transmit any bit. The channel state transition probabilities are described by

$$p_{ij} = p(h_t = i \,|\, h_{t-1} = j), \text{ with } i, j \in \{0, 1\}. \qquad (1)$$

These probabilities are assumed time-invariant and may be estimated using learning techniques [11]. here, they are assumed known *a priori*.

3.1.2 Buffers

The states of the l-th post-encoder and receiver buffer, with $l \in \{1 \dots\}$, are denoted by $s_l^e \in \mathcal{S}_l^e$ and $s_l^r \in \mathcal{S}_l^r$. They represent the level of the corresponding buffer. The vectors of states of all post-encoder and receiver buffers are respectively denoted by $\mathbf{s}_t^e = (s_{1,t}^e, \dots, s_{L,t}^e)$ and $\mathbf{s}_t^r = (s_{1,t}^r, \dots, s_{L,t}^r)$.

Various granularity levels may be considered to represent the content of a buffer [7, 8]. To minimize complexity, a coarse representation of the levels of the buffers is considered. Since buffer under and overflow have to be avoided, the values taken by the levels are quantized to get $\mathcal{S}_l^x = \{1, 2, 3\}$ with $x \in \{e, r\}$, where 1 represents underflow, 3 overflow, and 2 satisfying level.

3.2 Actions

The proxy has to determine the number of packets from each layer to send. When the channel conditions are bad and to avoid post-encoder buffer overflow, packets may also be dropped. The action $a_{l,t}$ taken for the l-th layer at time t represents then the number of transmitted packets from the post-encoder buffer, when its value is positive, or the number of dropped packets when it is negative. If $a_{l,t} = 0$, packets are neither transmitted nor dropped. The vector gathering all actions is denoted by $\mathbf{a} = (a_1, \dots, a_L) \in \mathcal{A}$.

3.3 Transition matrix and reward function

Once all states and actions have been identified, one has to determine the 3D transition probability matrix

$$T(\mathbf{s}_t, \mathbf{s}_{t+1}, \mathbf{a}_t) = \Pr(\mathbf{s}_{t+1} \,|\, \mathbf{s}_t, \mathbf{a}_t), \qquad (2)$$

with $\mathbf{s}_t = (\mathbf{s}_t^e, \mathbf{s}_t^r, h_t)$.

At time t, the proxy has to apply an action that maximizes the received video quality while satisfying the constraints described in Section 2.2. For that purpose, the following reward function is introduced.

$$\begin{aligned} R_t(\mathbf{s}_t, \mathbf{a}_t) \;=\; & \underbrace{\sum_{l=1}^{L} \gamma_l a_{l,t}}_{\text{transmission}} + \underbrace{\beta\nu\left(R_{T,t}(\mathbf{a}_t, h_t) - R_{c,t}\right)}_{\text{bandwidth constraint}} \\ & + E[\underbrace{\sum_{l=1}^{L} \lambda_l \rho(s_{l,t}^e, a_{l,t})}_{\text{encoder buffer}} + \underbrace{\sum_{l=1}^{L} \mu_l \rho(s_{l,t}^r, a_{l,t})}_{\text{receiver buffer}}]. \end{aligned} \qquad (3)$$

The positive parameters γ_l, λ_l, μ_l, with $l = 1 \dots L$, and β help to trade off the importance of the various constraints. The reward function (3) involves several parts, the first linked to the received video quality, the others to the constraints mentioned in Section 2.2.

Assuming that increasing the amount of transmitted packets increases the received quality, the transmission reward should help to maximize the amount of transmitted packets. The parameters γ_l allow to give a higher priority to packets belonging to the base layer compared to those of the enhancement layers.

For encoder and receiver buffer constraints, $\rho(.)$ provides a positive reward for buffer State 2 and a negative reward for States 1 and 3.

$R_{T,t}(\mathbf{a}_t, h_t)$ is the total transmission rate at time t when the action is $\mathbf{a}_t$ and the channel state is h_t. The function $\nu(x) = -\text{abs}(x) - \nu_0 \text{sgn}(x)$ introduces a strong penalty when $R_{T,t}(\mathbf{a}_t, h_t) - R_{c,t} > 0$, corresponding to a required rate larger than the actual channel transmission rate. When $R_{T,t}(\mathbf{a}_t, h_t) - R_{c,t} \le 0$, a positive reward is provided, maximum when the difference vanishes.

3.4 Myopic and foresighted policies

Unlike traditional *myopic* rate control optimization techniwues, which focuses on the maximization of some immediate reward, the goal in the proposed rate control framework is to find an optimal policy for each SNR layer that maximizes the expected discounted sum of future rewards

$$\sum_{t=0}^{\infty} \alpha^t R_t(\mathbf{s}_t, \mathbf{a}_t | \mathbf{s}_0). \qquad (4)$$

where the parameter $0 < \alpha < 1$ is the discount factor, which defines the relative importance of present and future rewards, and $\mathbf{s}_0$ is the initial state.

The *foresighted* policy, which maximizes the above sum when $\alpha > 0$, takes into account the impact of the current actions on the future rewards. When $\alpha = 0$, only the immediate reward is maximized and the corresponding policy is called *myopic*.

Both policies are obtained by value iteration as detailed, *e.g.*, in [11].

4. EXPERIMENTAL RESULTS

The performance of the proposed layer filtering process has been evaluated on various video sequences (Foreman, Mother & daughter,...). Here the results for Foreman in QCIF format at $f_r = 30$ fps are reported. Similar results are observed for the other sequences. Experiments are performed using the H.264/SVC encoder.

The Foreman sequence is encoded using three MGS scalability layers per frame ($L = 3$) corresponding to cumulated average rates (PSNR for luminance) of 34.7 kbits/s (28.67 dB) for Layer 1, 107.0 kbits/s (31.5 dB) for Layer 1 and 2, and 327.0 kbits/s (35.82 dB) for all layers. The channel rate in its good state is $R_c = 300$ kbit/s. The channel state transition probabilities are $p_{11} = 0.9$ and $p_{00} = 0.8$, resulting in an average channel rate of 200 kbits/s. Four possible actions per layer are considered at each time instant $\mathcal{A} = \{-1, 0, 1, 2\}$.

The post-encoder and the receiver buffers are assumed to have a maximum size (in term of number of packets) $S^e = 20$ and $S^r = 30$. The levels at which they are considered in over and underflow are $S^e_{max} = 19$ and $S^e_{min} = 6$ for the post-encode buffers. The underflow limit for the receiver buffers is $S^r_{min} = 13$.

The values of the parameters in the reward function (3) reflect the importance of the various constraints. Some training provides $\gamma_1 = 300$, $\gamma_2 = 150$, $\gamma_3 = 60$, $\lambda_1 = 200$, $\lambda_2 = 100$, $\lambda_3 = 40$, $\mu_1 = 300$, $\mu_2 = 150$, and $\mu_3 = 60$. The parameters β and ν_0 of the bandwidth constraint are respectively set to 0.1 and 5000 to give more weight to the bandwidth constraint compared to other constraints. AAA Mihaela, do you have any good way to perform the tuning apart from doing it by hand as was done by Nesrine ? ZZZ

Performances obtained with a myopic policy are compared to those obtained with a foresighted policy with $\alpha = 0.9$.

The evolution of the PSNR for the luminance of the decoded video streams for both strategies are represented in Figure 2.

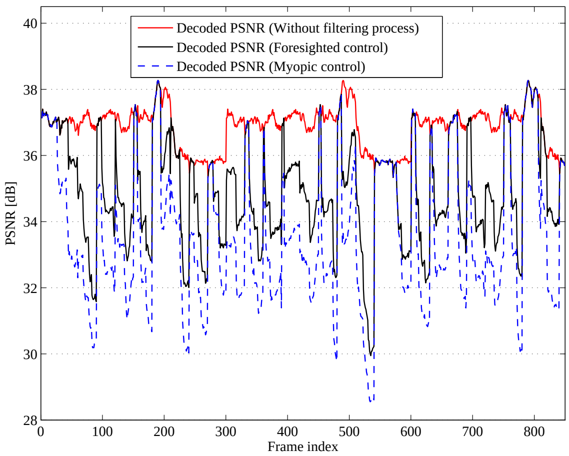

Figure 2: PSNR of the decoded sequence

AAA Nesrine, provide a plot for an other video sequence. Adapt the comments below ZZZ

In average, a gain of about 1.36 dB is obtained with the foresighted policy compared to the myopic one. This gain is mainly due to more packets of the first enhancement layer reaching the receiver. With the foresighted policy, 20.3% of the actions for the second enhancement layer are drop actions compared to 20.9% in the case of a myopic policy. For what concerns the first enhancement layer, no drop actions are obtained by using the foresighted policy compared to 20.3% for the myopic one.

An analysis of the level of the receiver buffers shows that they are more often at a satisfying level with the foresighted policy than with the myopic policy AAA Nesrine, give percentage of the time in both cases)ZZZ. This allows a better playback margin to be obtained with the foresighted policy.

The impact of controlling the level of the receiver buffers is easily evaluated by setting $\mu_l = 0$, $l = 1 \ldots 3$. AAA Nesrine, provide a plot, or provide numbers, An increase of the video quality when receiver buffer information is exploited.ZZZ

5. CONCLUSIONS AND PERSPECTIVES

This paper presents a scalable video streaming system over a time-varying wireless channel to a mobile user cast in the framework of MDP. Experimental results illustrate an improvement in the average PSNR with the foresighted policy compared to myopic policy. Considering receiver buffers contributes in some video quality improvements.

The proposed filtering process performs for video-on-demand services. For real-time video transmission, the parameters of the off-line control can be adapted to the variation of the video characteristics using reinforcement learning designed for on-line learning MDP policies [11]. Additional test benchmark, such as that in [6], will be considered.

6. REFERENCES

[1] J. Cabrera, A. Ortega, and J. Ronda. Stochastic rate-control of video coders for wireless channels. *IEEE trans. Circ. Syst. for Vid. Techn.*, 12(6):496–510, 2002.

[2] H. Chi-Yuan, A. Ortega, and M. Khansari. Rate control for robust video transmission over burst-error wireless channels. *IEEE Jnl Sel. Areas in Comm.*, 17(5):756–773, 1999.

[3] P. de Cuetos and K. W. Ross. Optimal streaming of layered video: joint scheduling and error concealment. In *ACM Multimedia*, pages 2–8, Berkeley, 2003.

[4] M. Hassan, L. Atzori, and M. Krunz. Video transport over wireless channels: a cycle-based approach for rate control. In *ACM Multimedia*, pages 916–923, New-York, 2004.

[5] M. Kalman, E. Steinbach, and B. Girod. Adaptive media playout for low-delay video streaming over error-prone channels. *IEEE trans. Circ. Syst. for Vid. Techn.*, 14(6):841 – 851, 2004.

[6] G. Liebl, H. Jenkac, T. Stockhammer, and C. Buchner. Radio link buffer management and scheduling for wireless video streaming. *Telecomm. Syst.*, 30(1-3):255–277, 2003.

[7] N. Mastronarde and M. van der Schaar. Designing autonomous layered video coders. *Sig. Proc. Image Comm.*, 24(6):417–436, 2009.

[8] N. Mastronarde and M. van der Schaar. Online reinforcement learning for dynamic multimedia systems. *IEEE trans. Image Proc.*, 19(2):290–305, 2010.

[9] P. V. Pahalawatta and A. K. Katsaggelos. Review of content-aware resource allocation schemes for video streaming over wireless networks. *Wireless Comm. and Mobile Comp.*, 7:131–142, 2007.

[10] H. Schwarz, D. Marpe, and T. Wiegand. Overview of the scalable video coding extension of the H.264/AVC standard. *IEEE trans. Circ. Syst. for Vid. Techn.*, 17(9):1103–1120, 2007.

[11] R. Sutton and A. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 1998.